PATENT ABSTRACTS OF JAPAN

(11)Publication number:

2002-132811

(43)Date of publication of application: 10.05.2002

(51)Int.CI.

G06F 17/30 G06F 17/27

G06F 17/28

(21)Application number: 2000-319998

(71)Applicant: NIPPON TELEGR & TELEPH CORP

<NTT>

(22)Date of filing:

19.10.2000

(72)Inventor: SASAKI YUTAKA ISOZAKI HIDEKI

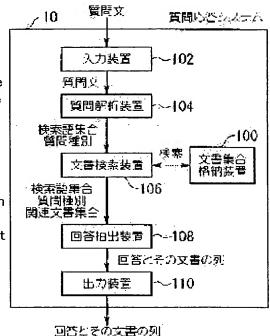
TAIRA HIROYORI KAZAWA HIDETO HIROTA KEIICHI NAKAJIMA HIROYUKI HIRAO TSUTOMU KATO TSUNEAKI

(54) METHOD AND SYSTEM FOR ANSWERING QUESTION AND RECORDING MEDIUM WITH RECORDED QUESTION ANSWERING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To output an answer to a question sentence and its document when a document collection is given.

SOLUTION: This question answering system outputs the answer to the question sentence and its document once given the document collection and question sentence. The system has a document collection storage device 100 which stores document collections, an input device 102 which receives the question sentence, a question analyzing device 104 which decides a retrieval word collection and the kind of the question from the question sentence obtained from an input device, a document retrieving device 106 which retrieves a related document collection from the document collections stored in the document collection storage devices, an answer extracting device 108 which extracts answers to the question sentence from respective documents in the related document collection and generates an array of the answers and the documents having the answers



extracted as an answer result to the question sentence, and an output device 110 which outputs the answer result.

LEGAL STATUS

[Date of request for examination]

21.12.2001

[Date of sending the examiner's decision of rejection

[Kind of final disposal of application other than

the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

e.

€ 翐 4 盂 华 噩 (<u>2</u>) (18) 日本国格群庁 (JP)

特開2002-13281

(11)特許出關公開番号

(P2002-132811A)

(43)公開日 平成14年5月10日(2002.5.10)

(51) Int CL.		說別記号	FI		本	テーマコード(参考)
	17/30	330	G 0 6 F 17/30		330C	6B076
		170		-	170A	5B091
	12/11		//1	22	ш	
	17/28		11/28	88	۲	

開求項の数9 OL (全 10 頁) 医左后腔 人名

(71) 出版人 000004226 日本電信電路株式会社	東京都千代田区大手町二丁目3番1号 佐々木 裕 東京都千代田区大手町二丁目3番1号 日	本電信電話株式会社内機局所有的工作。 不可以 化二甲基酚 化二甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲基酚甲	(74)代型人 100064308 弁理士 志賀 正氏	一般体質に続く
(71) 出版人	(72)発明者	(72) 発明者	(74)代理人	
徐昭 2000-319998(P2000-319988)	平成12年10月19日(2000.10.19)			
(21) 出原番号	(22) 掛鼠日			

質問な答方法、質問応答システム及び質問応答プログラムを配録した記録媒体 (34) [現形の名集]

[膜題] 文書集合が与えられると質問文に対する回答 とその文書を出力できるようにする。

【解決手段】 文書集合と賢問文が与えられると、賢問 文に対する回答と文書の列を出力する質問応答システム であって、文書集合を格納する文書集合格約装置100 と、質問文を受けとる入力装置102と、入力装置から 得られた質問文から検索語集合と質問種別を判定する貿 問解析装置104と、前記後索語集合と質問種別に従っ て、文書集合格納装置に格納された文書集合から関連文 野集合を検索する文書検索装置106と、関連文書集合 中の各文部から質問文に対する回答を抽出し、該回答と 核回答を抽出した文書の列を質問文に対する応答結果と して作成する回答抽出装置108と、応答結果を出力す る出力核電110とを有する。

質問応答システム 文智集合格的装置 回答とその文色の列 <u>=</u> 回衛苗出城區 108 <u>~</u> 7 質問解析義置 文書検索装置 出力被留 入力被配 検索簡集合 質問確別 関連文語集合 質問文 유

題文に対する回答と文章の列を出力するコンピュータシ 【請求項1】 文書集合と質問文が与えられると、該質 ステムを使用した質問応答方法であって、

た文書集合から関連文書集合を検索し、訪関連文書集合 女響の列を前記質問文に対する応答結果として出力する 抜検索語集合および該質問種別に従って、前記与えられ の各文書から回答を抽出し、該回答と該回答を抽出した 入力された質問文から検索語集合と質問種別を判定し、 ことを特徴とする質問応答方法。

【請求項2】 前記閣選文書集合の要素は文書全体では なく、文書の一部として前配文書集合から前記関連文書 集合を検索することを特徴とする請求項1に記載の質問

とを特徴とする請求項1または2のいずれかに記載の質 【請求項3】 前記閩漢文書集合を検索する際に計算し た各文書のスコアである文書スコアと前配関連文書集合 の各文書から回答を抽出する際に計算した抽出スコアの 2つのスコアに従って、回答と文書の列を頤序付けるこ 間応答方法。

徴とする請求項1乃至3のいずれかに記載の質問応答方 【請求項4】 前記関連文書集の各文書から回答を抽出 する際に、固有名詞や数値表現の認識を行なうことを特

【精求項5】 文書集合と質問文が与えられると、該質 問文に対する回答と文書の列を出力する質問応答システ

文書集合を格納する文書集合格制装置と、

抜入力装置から得られた前記質問文から検索語集合と質 質問文を受けとる入力装置と、

前記検索語集合と質問禮別に従って、前記文書集合格的 装置に格納された文書集合から関連文書集合を検索する 問種別を判定する質問解析装置と、

関連文書集合中の各文書から質問文に対する回答を抽出 し、飯回答と駿回答を抽出した文書の列を前記覧間文に 対する応答結果として作成する回答抽出装置と、

文書検索装置と、

を有することを特徴とする質問応答システム。 前記応答結果を出力する出力装置と、

素を文書全体ではなく、文書の一部として前配文書集合 【請求項6】 前記文事検索装置は、関連文書集合の要 から前記関連文書集合を検索することを特徴とする請求 頃5に記載の質問応答システム。

である文書スコアと前記回答袖出装置が前記関連文書類 の2つのスコアに従って、回答と文章の列を順序付ける ことを特徴とする間水項5または6のいずれかに記載の 【精水項7】 | 前記回答抽出裝置は、前配文書検索装置 が関連文書集台を検索する際に計算した各文書のスコア 台の各文書から回答を抽出する際に計算した抽出スコア

回答とそが文書の列

の各文書から回答を抽出する際に、固有名詞や数値表現 の認識を行なうことを特徴とする請求項5乃至7のいず [請求項8] 前記回答抽出裝賦は、前記関連文卷集合 れかに配載の質問応答システム。

【腈水項9】 文書集合と質問文が与えられると、該質 間文に対する回答と文魯の列を出力する質問応答を行う ための質問応答プログラムを記録したコンピュータ読み 取り可能な記録媒体において、

入力された質問文から検索語集合と質問種別を判定する 質問文を受け取る第1のステップと、 斑2のストップと、

た文書集合から関連文書集合を検索する第3のステップ 前記検緊語集合および質問種別に従って、前記与えられ

前記関連文書集合の各文書から回答を抽出し、歓回答と 抜回答を抽出した文書の列を作成する第4のステップ 前配回答と数回答を抽出した文書の列を前記質問文に対 **ドコンピュータに実行させる質問応答プログラムを記録** する広答結果として出力する第5のステップと、 した記録媒体。

[発明の詳細な説明]

[000]

(発明の属する技術分野) 本発明は、自然昏語処理シス テム、倉籍処理システム、知識処理システム、情報検索 システム、情報抽出システム等に用いられ、質問文に対 する回答を出力する質問応答方法、質問応答システム及 が質問応答プログラムを記録した記録媒体に関する。

集合から、ユーザの入力した質問に合致する文書の集合 められた項目を抽出するものであった。従来の日本語質 ていたが、その回答の元となる記事を同時に出力してい 【従来の技術】従来の情報検索技術は、与えられた文書 は、与えられた1つの文書について、分野ごとに予め決 問応答システムは、回答として単語や単語の列を出力し を取り出すものであった。また、従来の情報抽出技術 [0002]

ガスコアを総合的に用いて、質問に対して出力する回答 と文書の列の順序を決める等の、回答の精度を向上させ 【0003】さらに、固有名詞の抽出技術の利用や、質 の文書スコア、及び回答を抽出する際の抽出スコアの両 問文に対する回答を含む可能性のある文書を検索する際 るための工夫が行なわれていなかった。

「日本の首相は誰ですが?」という質問に対して、返っ [発明が解決しようとする課題] まず、従来の情報検索 技術は質問に対する結果を文書の単位で応答するため、 ューザが文書の内容を抵まなければならない。例えば、 てきた結果に含まれる文書を読むことにより初めて、

「森首相」という答が判る。文書を読まなければならな

3

いことは「禁首相」といった直接的な回答が欲しいユー **ザにとっては非常に煩わしいという問題が有った。**

【0005】次に、従来の情報抽出技術は、予め決めら する回答を出力する質問応答システムでは使えない。従 来の日本語覧問応答システムは、回答として単語や単語 の列を出力していたが、回答の元となる記事を返してい ができながった。例えば、「ジンパブエの大統領は誰で すか?」という質問に対して、「ムガベ大統領」と回答 が返ってきても、ユーザが本当に「ムガベ大航領」が正 れた特定の分野でしか使えないため、任意の質問文に対 なかったため、回答を裏付ける情報をユーザが得ること しいがどうかを確認できないという問題が有った。

ものであり、質問文に対する直接的な回答とその回答の 元となる文書を提示することにより回答の信頼性をユー ザが確認することができる質問応答方法、質問応答シス テム及び質問応答プログラムを記録した記録媒体を提供 **【0006】本独明はこのような事情に貼みてなされた** することを目的とする。

[0000]

【課題を解決するための手段】上記目的を達成するため られると、眩賢問文に対する回答と文書の列を出力する て、入力された質問文から検索語集合と質問確別を判定 し、駄検索語集合および舷質問種別に従って、前配与え 集合の各文書から回答を抽出し、該回答と該回答を抽出 した文書の列を前記質問文に対する応答結果として出力 に、臍水項1に記載の発明は、文書集合と質問文が与え られた文書棋合から関連文書集合を検索し、鼓関連文書 コンピュータシステムを使用した質問応答方法であっ することを特徴とする。

た質問文から検索語集合と質問種別を判定し、鼓検索語 [0008] 請求項1に記載の発明によれば、入力され 集合および眩覚問種別に従って、前記与えられた文書集 合から関連文書集合を検索し、鼓関連文書集合の各文書 から回答を抽出し、該回答と該回答を抽出した文書の列 を前記質問文に対する応答結果として出力するようにし なる文哲を提示することにより回答の信頼性をユーザが たので、質問文に対する直接的な回答とその回答の元と 確認することができる。

[0009]また請求項2に記載の発明は、請求項1に 記載の質問応答方法において、前記関連文書集合の要素 は文書全体ではなく、文書の一部として前記文書集合か 【0010】 請求項2に記載の発明によれば、請求項1 に記載の質問応答方法において、前記関連文書集合の要 紫は文書金体ではなく、文書の一部として前配文書集合 ら前記関連文書集合を検索することを特徴とする。

[0011] また、請求項3に記載の発明は、請求項1 または2のいずれかに記載の質問応答方法において、前 記関連文書集台を検索する際に計算した各文書のスコア 的な回答が得られる。

から前記関連文書集合を検索するようにしたので、直接

または2のいずれかに記載の質問応答方法において、前 である文書スコアと前記関連文書集合の各文書から回答 である文書スコアと前記関連文書集合の各文書から回答 を抽出する際に計算した抽出スコアの2つのスコアに従 記聞連文書集合を検索する際に計算した各文書のスコア を抽出する際に計算した抽出スコアの2つのスコアに従 って、回答と文書の列を顧序付けるようにしたので、賢 [0012] 請求項3に配載の発明によれば、請求項1 って、回答と文書の列を順序付けることを特徴とする。 問文に対する回答の精度の向上が図れる。

乃至3のいずれかに記載の質問応答方法において、前記 関連文書集の各文書から回答を抽出する際に、固有名詞 [0013] また、請求項4に記載の発明は、請求項1 や数値表現の認識を行なうことを特徴とする。

乃至3のいずれかに記載の質問応答方法において、前記 関連文容集の各文書から回答を抽出する際に、固有名詞 [0014] 請求項4に記載の発明によれば、請求項1 や数値表現の認識を行なうようにしたので、質問文に対 する精度の向上が図れる。

[0015]また、請求項5に記載の発明は、文書集合 と質問文が与えられると、該質問文に対する回答と文書 格納する文書集合格納装置と、質問文を受けとる入力装 置と、抜入力装置から得られた前配質問文から検索語集 合と質問種別を判定する質問解析装置と、前記検索語集 合と質問種別に従って、前記文審集合格納装置に格納さ れた文書集合から関連文書集合を検索する文書検索装置 と、関連文書集合中の各文書から質問文に対する回答を 抽出し、玆回答と該回答を抽出した文書の列を前記賢問 文に対する応答結果として作成する回答抽出装置と、前 兄広答結果を出力する出力装置とを有することを特徴と の列を出力する質問応答システムであって、文書集合を \$ 50

[0016] 請求項5に記載の発明によれば、文書集合 を格納する文書集合格納茲置と、質問文を受けとる入力 装置と、抜入力装置から得られた前配貨間文から検索語 集合と賢問禮別を判定する質問解析装置と、前配檢索語 集合と質問種別に従って、前記文曹集合格納装置に格約 された文書集合から関連文書集合を検索する文書検索装 置と、関連文普集合中の各文書から質問文に対する回答 前記広答結果を出力する出力装置とを有するので、質問 文に対する直接的な回答とその回答の元となる文書を提 を抽出し、該回答と該回答を抽出した文書の列を前記賞 **示することにより回答の信頼性をユーザが確認すること** 問文に対する広答結果として作成する回答抽出装置と、

[0017] また、請求項6に記載の発明は、請求項5 に記載の質問応答システムにおいて、前記文書検索装置 は、関連文書集合の要素を文書全体ではなく、文書の一 部として前記文書集合から前記関連文書集合を検索する ことを特徴とする。

に記載の質問応答システムにおいて、前配文舎検索装置 [0018] 請求項6に記載の発明によれば、請求項5 都として前配文書集合から前配関連文書集合を検索する は、関連文書集合の要素を文書全体ではなく、文書の一 ので、直接的な回答が得られる。

【0019】また、請求項7に記載の発明は、請求項5 て、前記回答抽出装置は、前記文書後索装置が関連文書 集合を検索する際に計算した各文書のスコアである文書 スコアと前記回答抽出装置が前記関連文書集合の各文書 から回答を抽出する際に計算した抽出スコアの2つのス コアに従って、回答と文書の列を順序付けることを特徴 または6のいずれかに記載の質問応答システムにおい

コアに従って、回答と文書の列を順序付けるようにした 【ロの20】請求項フに記載の発明によれば、請求項5 て、前記回答抽出裝置は、前記文書検索装置が関連文書 集合を検索する際に計算した各文書のスコアである文書 スコアと前記回答抽出装置が前記関連文書集合の各文書 から回答を抽出する際に計算した抽出スコアの2つのス または6のいずれかに記載の質問応答システムにおい ので、質問文に対する回答の精度の向上が図れる。

答を抽出する際に、固有名詞や數値表現の認識を行なう 【0021】また、請求項8に記載の発明は、請求項5 前記回答抽出装置は、前記関連文書集合の各文書から回 乃至うのいずれかに記載の質問応答システムにおいて、 ことを特徴とする。

答を抽出する際に、固有名詞や数値扱現の認識を行なう [0022] 請求項8に記載の発明によれば、請求項5 前記回答抽出装置は、前記関連文書集合の各文書から回 乃至7のいずれかに記載の質問応答システムにおいて、 ので、質問文に対する回答の精度の向上が図れる。

[0023]また、請求項9に記載の発明は、文書集合 ムを記録したコンピュータ読み取り可能な記録媒体にお ップと、前記検索額集合および賢問種別に従って、前記 し、核回答と核回答を抽出した文書の列を作成する第4 の列を出力する質問応答を行うための質問応答プログラ いて、質問文を受け取る第1のステップと、入力された 質問文から検索語集合と質問種別を判定する第2のステ 与えられた文書集合から関連文書集合を検索する第3の ステップと、前記関連文書集合の各文書から回答を抽出 のステップと、前記回答と該回答を抽出した文書の列を をコンピュータに実行させる質問応答プログラムを記録 と質問文が与えられると、該質問文に対する回答と文書 前記貨間文に対する応答結果として出力するステップと した記録媒体を要旨とする。

[0024] 請求項9に記載の発明によれば、文書集合 ムを記録したコンピュータ読み取り可能な記録媒体にお いて、質問文を受け取る第1のステップと、入力された と質問文が与えられると、該質問文に対する回答と文書 の列を出力する質問応答を行うための質問応答プログラ

ップと、前記検索語集合および賢問権別に従って、前記 せ、実行することにより、質問文に対する直接的な回答 与えられた文智集合から関連文書集合を検索する第3の ステップと、前記関連文書集合の各文書から回答を抽出 ップとをコンピュータに実行させる質問応答プログラム を記録媒体に記録したので、この記録媒体に記録した質 とその回答の元となる文書を提示することにより回答の し、該回答と該回答を抽出した文書の列を作成する第4 のステップと、前記回答と該回答を抽出した文書の列を 竹記質問文に対する応答結果として出力する饼5のステ 質問文から検索語集合と質問種別を判定する第2のステ 間応答ブログラムをコンピュータシステムに続み込ま 宮頼住をユーザが確認することができる。

の形態に係る質問応答システムは、質問文に対する直接 取り出した文書をユーザに出力することにより、回答の 【発明の実施の形態】以下、本発明の実施の形態を、図 面を参照して詳細に説明する。図1に本発明の実施の形 想に係る質問応答システムの構成を示す。 本発明の実施 的な回答として文磬の一郎を取り出すとともに、回答を **宮頼性をユーザが確認できるようにしている。**

[0025]

および関連文書集合から回答を抽出する際における抽出 する回答と文書の列の頌序を決めることにより、質問に 対する正しい回答を回答と文書の列のより上位に並べる [0026]また、固有名詞の抽出技術や、与えられた 文書集合から関連文書を検索する際における文書スコア スコアの西方を総合的に用いて評価することより、出力

合および鼓質問種別に従って、前記与えられた文都集合 ら回答を抽出し、駄回答と馼回答を抽出した文書の列を [0027] すなわち、本発明の実施の形態に係る質問 **応答システムは、文書集合と質問文が与えられると、該** から関連文書集合を検索し、拡関連文書集合の各文書か 前記質問文に対する応答結果として出力することを特徴 システムを使用した質問応答方法であって、入力された 質問文から検索語集合と質問種別を判定し、該検索語集 質問文に対する回答と文巻の列を出力するコンピュータ ことを可能にしている。

[0028]図1において、本発明に係る質問応答シス 質問文を解析し、上記質問文から核索語集合と質問機別 別に従って文書集合格納装置100に格納されている文 **哲集台から関連文音集合を検索する文書検索装置106** し、該回答と該回答を適出した文眷の列を上記質問文に テム10は、与えられた文書集合を格納する文書集合格 を判定する質問解析装置104と、検索語集合と質問種 対する応答結果として作成する回答抽出装置108と、 納装置100と、質問文を受け取る入力装置102と、 と、関連文書集合から上記質問文に対する回答を抽出 上記広答結果を出力する出力装置110とを有してい とする質問応答方法を実施するための装置である。

(9)

トに基づいて説明する。まず、文書集合格納装置100 1)。 質問解析装置 104は質問文に対して形態素解析 [0029] 上記構成からなる本実施の形態に係る質問 に、与えられた文書の集合が格納される(ステップ20 0)。入力装置102は、質問文が入力されると、その を行い、質問文から検索語集合と質問種別を判定し、検 **巧答システム10の処理内容を図2に示すフローチャー** 質問文を質問解析装置104に選す(ステップ20

から受け取った検索語集合と質問種別に従って、文書集 集合を抽出し、鼓関連文書集合の各文書に対して後索語 関連文書集合を、検索語集合及び質問種別と共に回答曲 索語集合と質問権別を文書検索装置106に渡す(ステ 合格約装置2に格約された文書集合を検索し、関連文書 が含まれている度合いを示す文書スコアを算出し、上記 【0030】文書検索装置106は質問解析装置104 ップ202)。

208),

【0031】回答抽出装置108は文容後需装置106 従った単語を抽出対象とし、抽出すると共に (ステップ から受け取った関連文書集合中の各文書について形態条 解析を行う(ステップ204)。そして回答抽出裝置1 0.8 は、形態紫解析を行った各文書に対して質問種別に 205)、上記抽出対象についてその抽出対象が合まれ

出数回108へ譲ず (ステップ203)。

る文色中における検索語との距離に基づいて抽出スコア

	文響番号	.
	10	「日本の茶首相とグリトン・アメリカ大統領が会談した
	05	「ロシアのブーチン大統領が来日した」
	60	「昨日,京都で祇園祭が行なわれた」
•		

でもよい。また、質問文を1つ受けて、回答とその文書 本発明の実施の形態では、説明を容易にするため、文書 を1文だけで記述しているが、複数の文母からなる文書 の列を返す例を述べているが、これを繰り返すことによ り、質問と回答を繰り返すことができる。

1、D2、D3を格納する。ここでは、殺形式で表現し ているが、格納の方法は文告が格納できればリストやデ 【0035】まず、文哲集合格約装置100に文書D **ータベースなどの街の方法であっても何でもよい。**以 下、 数形式でデータを扱した時には同様のことが含え [0036]入力核配102は、質問文を受けとり、質 の名前を聞くwhatや、方法を聞くhowなどの質問 日時を聞くwhenの3種類とする。なお、この他の物 **陸別があったとしても、同様な方法で質問種別を判定す** 質問種別は人名を聞くwho.塩所を聞くwhere. は、質問文の質問種別を判定する。本英語の形態では、 間解析装置104~渡す。まず、質問解析装置104 ることができる。

を算出する (ステップ206)。

[0032]次いで、回答抽出装置108は、文書技索 装置106で算出した文書スコアと上配抽出スコアに従 回答を抽出し、この回答と、この回答を抽出した元とな る文書の列を取り出し、回答及び文書の列の頤序付けを 行う(207)。そして回答抽出装置108は、頤序付 けを行った回答及び文書を示す文書番号の列を応答結果 その文書番号の列を応答結果として出力する(ステップ って、上記関連文書集合中の各文書から質問文に対する として出力装置110~渡す。出力装置110は回答と

[0033]なお、本英施の形態に係る質問応答システ 各核関が1 台のコンピュータの上で通信しながら動く独 ワークで接続された複数のコンピュータに分散されて実 立したブロセスとして횢抜されていても良いし、ネット ムを構成する各装置は鎮理的につながっていればよく、 装されていても良い。

る例を述べる。まず、文書集合が扱1に示すように、3 [0034]次に、本発明の実施の形態に係る質問応答 システムの具体的動作について説明する。以下では、例 に対して、回答「クリントン」と記事番号「D1」を得 として、質問文ロı「アメリカの大統領は誰ですか?」 つの文香ひ1、ひ2、ひ3からなるとする。

[報1]

[0037] 質問種別の判定には、 扱2に示すような判

[表2]

質問種別の判定表現

定要現を持つ判定要を用いる。なお、判定要現は質問の

種別を分けるために使える語であれば、他の語でもよ

質問種別 判定蔑現	頂
who だれ。誰、どの人	<i>'</i> ፍወץ
where どこで	どこで、何処、どの国、国はどこ、何頼
when L1つ,倒	いつ,何時,何日,何月,何年,何曜日

し、判定数現が含まれる質問種別をG-1の質問種別GT とする。 G1には「雄」が含まれるので、G1の質問種 **判定表現が質問文Q1に含まれるかどうかをチェック** 路はかわっとなる。

折し、単語に分けるとともに、品詞の情報を得る。形態 [0038] 質問解析装置 104は、質問文を形態兼解 素解析の手法は例えば文献 (長尾真編:自然宮語処理、

各単語の品間のうち、前後の単語の関係から最適な品詞 文を辞書にある単語の列に分割し、辞書に書かれている は、辞書に含まれる単語が文に現れるかどうかを如べ、 岩波香店、1996)に述べられている。形態素解析 を選択することにより実現される。

[0039] 現在の例では、「アメリカの大統領は誰で すか?」は、数3のように形態素解析されるとする。 [泰3]

01の形態素解析結果

88	固有名詞	格助酮	普通名詞	经体值	化名詞		國仰聲	52
単語	ተረነአ	6	大紙僧	12	奲	£2	ıΩ	b
単語番号	1	2	8	4	5	9	7	8

は、説明の簡単化と検索精度の向上を図るためであるの で、すべての単語を検索語集合にしたり、他の選択法に [0041] 質問解析装置104は後案語集合KWと買 と豊ける。ここで、自立語のみを検索語集合とするの よって検索語集台を選択したりしてもかまわない。

曹集合格的装置 100に格納されている文書集合から探 (文書番号で扱される。) の文書スコアとする。この結 問種別のTを文書検索装置106に渡す。文書検索装置 106は、後索語集合KWがより多く合まれる文書を文 す。検索語集合の含まれている数を数え、それを各文書 果を喪4に示す。

文書207

RD= (D1, D2)

となる。なお、文書スコアの計算法はIDF法やTFI もよい。1DF法やTF1DF法の計算式や検索の高速 化のためのインデックス作成法は、例えば、文献(徳永 健伸:情報検索と倉語処理、東京大学出版会、1999 DF法など、当業界において用いられる方法ならなんで 年)に述べられている。

[0044] また、ここでは文書スコアの計算には、各 文書中に検索語集合の要素が現れるかどうかを文書中の

なお、ここで品詞名は宜語学や自然宮語処理において利 用されるものであれば何でもよい。例えば、普通名詞を 一般名詞と毀現してもよい。このうち、普通名詞、固有 名詞などの自立語を検索語集合KWとする。

[0040] 上記質問文から検索語集合KWは、

らに、文書番号D3の文書は、検索語集合の要素を含ま 文書番号D1の文書の文書スコアは2となる。また、文 「大統領」しか含まないので文書スコアは1となる。さ 「大統領」の2つの検索脂集合の要素が含まれるので、 数番号ひ2の文書スコプは、核索語集合の要索である 【0042】文書番号D1の文書には「アメリカ」、 ないので文魯スコアは0である。 **【0043】文告検索装団106は文書スコアが0より** 大きい文哲を関連文書集合RDとする。したがって、

の単位としてもよい。文告後索装置106は関連文書集 合RDを、検索語集合及び質問種別と共に、回答抽出装 問108に渡す。なお、関連文書集合RDは文書全体を の単位としたり、100文字といったパッセージを検索 文字で調べたが、文書を形態素解析し、単語に分けてか ら文書中の単語と検索語を比較してもよい。 さらに、文 哲全体を1まとまりとしなくても、文音の各段落を検索 液しても良いし、文書の名前だけを滾してもよい。 (8)

【0045】回答抽出较配108は、関連文學集合RD 中の文在を形態素解析する。文書番号D1と文書番号D 2の各文書の形態繁解析結果をそれぞれ数5、数6に示

語がく人名>、〈地名〉、〈日時〉という3種類の固有 【0046】次に固有扱現抽出法により、文書中の各単 **扱現のどれかに拡当するかどうかを判定し、拡当する場** 合はその種別を単語に付与する。基本的には固有名詞の 辞費と周囲の単語により人名、地名、日時であるかを判 定する。固有殺現の抽出法については、例えば特開平1 1 - 0 6 7 5 6 2 に記倣されている。固有投現の判定後 の結果を扱て、張8に示す。

[0047]

文書01の形態繁解析結果 (報2)

_													
	89	國名吳国	國佣針	四百名詞	健多更具	國仰檀	國名其国	鲁23	國足量	國多更暴	随仰针	壁傳	财团和
	日日	≭ 8	ø	#	財算	7	CACIG	•	ሲኒሉፈ	大概領	JΩ	弾号	37
	单码路号	1	2	8	4	5	9	7	8	6	10	11	12

女母02の形態発解析結果 (報6)

1000	國安共團	格助阿	超多量图	普通名嗣	格助阿		即國阿
田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田	ብኝP	0	ブーチン	大統領	Δí	来日	ال الج
中医器号	. 1	2	Э	4	5	9	7

単語を抽出対象とする。whoの場合は品詞・固有費現 【0048】回衛抽出撤倒108は、敷配鞴別に抜った がく人名>である単語を抽出対象とし、whereの場 合は品詞・固有喪現がく場所>である単語を抽出対象と し、whenの場合は品駒・固有喪現が<日時>である

統領」と6単語離れているので単語「森」の抽出スコア

抽出対象に組織名を加えても良い。また、"「我蟹は猫 単語を抽出対象とする。但し、質問種別に対応する抽出 対象はこれに殴るわけではない。例えば、whereの である」。のようなカギ括弧で括られた部分を抽出対象 に加えても良い。

[0049]

[卷7]

文書のの固有表現抽出結果

品阿·固有表現	<地名>	格助酮	<大名>	普通名詞	自由自	<\$Y>	5 21	<志名>	普通名詞	格的阿	動胸	即即阿
超曲	B¥	9	H22	節相	L.	くれてばん		TUKE	大統領	Ę	敬	元
自器器号	-	2	6	4	5	9	7	80	٥	2		12

文像02の固有表現抽出結果

(報8)

品詞-固有表現	< 地名>	格助酮	7 〈人名〉	1 普通名阿	格助胸	. 配題	品を
四田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田田	G:P	9	プーチン	大統領	¥	#	14.
中四番号	-	2	Э	4	5	9	7

離れて出現するかを求めそれを距離とし、各要素につい [0051] 具体的には、単語間の距離は2つの単語の 単語番号の差の絶対値とする。文書番号D1の文書にお ける単語「森」は単語「アメリカ」と5単語、単語「大 [0050] 本実施の形態の場合は、質問種別がwho 「クリントン」、文書番号D2の文書における単語「ブ ここでは、抽出対象と検索語KWの各要素が何単語 **ーチン」が、抽出対象となる。 これらの抽出対象にしい** て、検索語KWとの距離を使って抽出スコアを計算す であるので、文哲番号ロ1の文書における単語「森」 て求められた距離の逆数の和を抽出スコアとする。

領」と3単語離れているので「クリントン」の抽出スコ は1/5+1/6=0.37となる。また、文書番号ロ1の アは1/2+1/3=0.83となる。さらに、文書番号D 1の文母における単語「ブーチン」は単語「大統領」と 1 単語離れているので単語「プーチン」の抽出スコアは 女告「クリントン」は「アメリカ」と2単語、「大統

[0052]次に、回答抽出装置108は、各文容中の

由出対象について、総合スコアを次のような総合計算式

1となる。なお、抽出スコアの計算は当象界で使われる

ものであれば、他のものでもよい。

総合スコア= a×文存スコア+ b×抽出スコア で来める。

但し、a、bはそれぞれ文啓スコア、抽出スコアの塩み

【0053】倒えば、a、b、c、dをパラメータとし **集界で用いられるものであれば他の方法でも掛わない。** を殺すパラメータであり様々な定め方がある。ここで

は、ョ=1、 b=1とする。なお、スコアの計算法は当

総合スコプ=a×文也スコア+b×段塔スコア+c×文スコア+d

×毎日スコア

のように多段化してもよい。ここで、文甞スコアの「文 文スコアの「文」は、読点までをひとかたまりとした文 字列の収合体である。各文督中の抽出対象についての抽 出スコア、文啓スコア、総合スコアを計算した結果を設 香」は、読点で区切られた複数の文容の集合体であり、 9 仁示寸。

[0054] [發9]

抽出対象のスコア

発放打集	毘	第8377	XEXEX	移台イング
概	٥	0.37	2	2.37
グリントン	5	0.83	2	2.83
プチン	22	ı	1	2

ム」とは、OSや周辺機器等のハードウェアを含むもの

[005:5] さらに、回答抽出敬酌108は、抽出対象 の総合スコアの大きい頃に抽出対象を回答とし、その文 笹番号を出力装置110に渡し、結果を出力する。ここ

- 1. 回答=「クリントン」、女母=D1 では、次ように出力される。
 - 2. 回答=「搡」、文母=D1
- 3. 回答=「プーチン」、文書=D2

でもよい。このようにして、文啓集合と「アメリカ大統 対する答えを文音集合から取り出し、回答とその回答を なお、出力の形式は同じ内容を殺していれば、他の形式 領は誰ですか?」という入力文が与えられると、それに **合む文書の列として出力された。**

ものも含むものとする。

[0056] 尚、図2に示す処理内容を質問応答プログ ラムとして作成し、このプログラムをコンピュータによ 用することによりコンピュータシステムに質問応答プロ グラムを実行させることにより質問応答システムの機能 り読み取り可能な記録媒体に記録し、この記録媒体を使 を実現するようにしてもよい。

[0057] すなわち、文書集合と質問文が与えられる と、舷質間文に対する回答と文書の列を出力する質問応 答を行うための質問応答ブログラムを記録したコンピュ 一夕読み取り可能な記録媒体において、質問文を受け取 る第1のステップと、入力された質問文から検索語集合

合および質問種別に従って、前配与えられた文曹集合か と質問稹別を判定する第2のステップと、前記検索語集 文書集合の各文書から回答を抽出し、歓回答と談回答を 抽出した文色の列を作成する第4のステップと、前配回 答と該回答を袖出した文書の列を前記質問文に対する応 答結果として出力する第5のステップとをコンピュータ この記録媒体に記録した質問応答プログラムをコンピュ 一タシステムに読み込ませ、契行することにより、質問 ら関連文告集合を検索する第3のステップと、前配関連 に実行させる質問広答ブログラムを記録媒体に記録し、 [0058] なお、ここでいう「コンピュータシステ **応答システムの機能を棄現するようにしてもよい。**

一タシステムに内型されるハードディスク等の記憶装置 媒体」とは、インターネット等のネットワークや電話回 **やクライアントとなるコンピュータシステム内部の原兜** 性メモリのように、一定時間プログラムを保持している 体」とは、フロッピー(登録商標)ディスク、光磁気デ 【0059】さらに「コンピュータ読み取り可能な記録 **镍等の通復回線を介してプログラムを送信する場合の通 엽線のように、短時間の間、動的にプログラムを保持す** るもの(伝送媒体ないしは伝送波)、その場合のサーバ **ィスク、ROM、CD-ROM等の可収媒体、コンピュ** とする。また、「コンピュータ読み取り可能な記録媒 のことをいう。

部を実現するためのものであっても良く、さらに前述し プログラムとの組み合わせで実現できるもの、所謂整分 [0060] また上記プログラムは、前述した機能の一 た機能をコンピュータシステムにすでに記録されている ファイル (差分プログラム) であっても良い。

に係る質問応答方法、質問応答システム及び質問応答ブ ログラムを配録した配録媒体によれば、自然官語で表現 また、本発明の実施の形態に係る質問応答方法、質問応 答システム及び質問応答プログラムを記録した記録媒体 [0061] 以上説明したように、本発明の実施の形態

によれば、質問文に対する直接的な回答とその回答のも された質問文に対する回答を求めることが可能となる。

サンローチャート。

[符号の説明]

特開2002-132811

6)

月用したりしていなかったため、回答の精度が低くなっ とになる文曲が求められるので、ユーザが直接回答を得 提示されるため、上記実施の形態において質問文の内容 である「大統領の名前」は、ユーザが文學を読んで初め て見つけることができたにすぎない。また、従来の情報 答システムは、回答ととも文鲁を提示していなかったの [0063] さらに、従来の質問応答システムでは、B 本語の処理に必要な単語分けを含む形態素解析を行なっ ていなかったり、固有扱現の抽出を行なっていなかった り、抽出対象のスコアを評価する際に検索時のスコアを 【0062】従来の情報検索技術では、文書全体の列が 抽出技術は分野に依存していたため、自由な質問文に対 する答えを抽出することはできなかった。従来の質問応 で、ユーザが回答の正しさを計ることができなかった。 るとともに、その回答の根拠を文掛でチェックできる。

は、このような問題を解決し、賢問文に対する回答とそ [0064]例えば、奥施例において、検索スコアを無 **現して、総合スコアを抽出スコアとすると、回答の列の 頃序が、「ブーチン」「クリントン」「森」の頭になっ** てしまう。また、形態素解析や固有扱現の抽出を行なわ ないと、柚出対象の単語が決まらず、回答の精度が低く なる。本発明の実施の形態に係る質問応答方法、質問応 答システム及び質問応答ブログラムを記録した記録媒体 の文書の列をユーザに提示できる効果がある。

[0065]

【発明の効果】以上に説明したように、請求項1に記載 の発明によれば、入力された質問文から後緊語集合と質 と該回答を抽出した文書の列を前記質問文に対する応答 接的な回答とその回答の元となる文書を提示することに 問題別を判定し、抜検索語集合および眩覚問種別に従っ **結果として出力するようにしたので、質問文に対する値** に記載の質問応答方法において、前記関連文書集合の要 **飛は文樹全体ではなく、文魯の一部として前記文都集合** から前記園迦文書集合を検索するようにしたので、直接 し、飯園迎文書集合の各文書から回答を抽出し、該回答 【0066】請求項2に記載の発明によれば、請求項1 て、前記与えられた文書集合から関連文書集合を検索 より回答の信頼性をユーザが確認することができる。

って、回答と文哲の列を順序付けるようにしたので、質 または2のいずれかに記載の質問応答方法において、前 である文香スコアと前記陸連文書集台の各文書から回答 5.抽出する際に計算した抽出スコアの2つのスコアに従 [0067] 情水頃3に記載の発明によれば、請水項1 **記関連文書集台を検索する際に計算した各文書のスコア** 間文に対する回答の推展の向上が図れる。

占な回答が律られる。

関連文書棋の各文書から回答を抽出する際に、固有名詞 [0068] 精水項4に記載の発明によれば、請水項1 5.英3のいずれかに記載の質問応答方法において、前記

や数値表現の認識を行なうようにしたので、質問文に対 する精度の向上が図れる。

【0069】請求項5に記載の発明によれば、文書集合 集合と質問種別に従って、前記文書集合格約装置に格約 された文書集合から関連文書集合を検索する文書検索装 置と、関連文書集合中の各文書から質問文に対する回答 を抽出し、該回答と該回答を抽出した文書の列を前記賞 文に対する直接的な回答とその回答の元となる文書を提 を格納する文書集合格納装置と、貿問文を受けとる入力 装置と、拡入力装置から得られた前配質間文から検索語 **集合と質問種別を判定する質問解析装置と、前記後索語** 前記広答結果を出力する出力装置とを有するので、質問 示することにより回答の信頼性をユーザが確認すること 問文に対する応答結果として作成する回答抽出装置と、 がてきる。

[0070] 請求項6に記載の発明によれば、請求項5 に記載の質問応答システムにおいて、前記文書検索装置 部として前記文書集合から前記関連文書集合を検索する は、関連文書集台の要素を文書全体ではなく、文書の一 ので、直接的な回答が得られる。

【0071】請求項フに記載の発明によれば、請求項5 て、前記回答抽出装置は、前記文書後紫装置が関連文書 集合を検索する際に計算した各文書のスコアである文書 スコアと前配回答抽出装置が前記関連文書集合の各文書 から回答を抽出する際に計算した抽出スコアの2つのス コアに従って、回答と文書の列を順序付けるようにした または6のいずれかに記載の質問応答システムにおい ので、質問文に対する回答の精度の向上が図れる。

前配回答抽出装置は、前配関連文書集合の各文書から回 項8に記載の発明によれば、請求項5乃至7のいずれか [0072] 請求項8に記載の発明によれば、請求項5 答を抽出する際に、固有名詞や数値表現の認識を行なう ので、質問文に対する回答の積度の向上が図れる。請求 **に記載の質問応答システムにおいて、前記回答抽出装置** こ、固有名詞や数値表現の認識を行なうので、質問文に 乃至7のいずれかに記載の質問応答システムにおいて、 は、前配関連文書集合の各文書から回答を抽出する際 対する回答の糖度の向上が図れる。

回答及び文書書号を出力

첫

のステップと、前記回答と該回答を抽出した文書の列を [0073] 請求項9に記載の発明によれば、文書集合 ムを記録したコンピュータ読み取り可能な記録媒体にお いて、質問文を受け取る第1のステップと、入力された **質問文から検索語集合と質問種別を判定する第2のステ** ップと、前記後祭語集合および質問種別に従って、前記 与えられた文書集合から関連文書集合を検索する第3の ステップと、前記関連文書集合の各文書から回答を抽出 、、飯回答と該回答を抽出した文書の列を作成する第4 前記質問文に対する広答結果として出力する第5のステ と質問文が与えられると、該質問文に対する回答と文書 の列を出力する質問応答を行うための質問応答プログラ

せ、実行することにより、質問文に対する直接的な回答 とその回答の元となる文書を提示することにより回答の を記録媒体に記録したので、この記録媒体に記録した質 ップとをコンピュータに実行させる質問応答プログラム 問応答プログラムをコンピュータシステムに読み込ま

[図1] 本発明の実施の形態に係る質問応答システム **間頼性をユーザが確認することができる。** 【図面の簡単な説明】

[図2] 図1に示す質問応答システムの処理内容を示 の構成を示すブロック図。

100 文色集合格格装配 10 質問応答システム 質問解析裝置 106 文色検索装置 回答抽出裝置 02 入力装置 出力装置 104 108

展局の高システム

四温文書集合を文書館合より接換文書を入口アを存出 文學集合反格的 ~200 毎田払表にしいて複数的との形態により複割として数数の 国選文書の形態素解析 質問文を入力 女馬女克尔克斯斯斯

文学スコア及び他出スコアに従って自答及び文書の別の記字付け 大學無力 回答とその文書の列 <u>=</u> 回街田田城村 ~108

回答とその文字の列

出力联盟

東京都千代田区大手町二丁目3番1号 本電信電話株式会社内 賀沢 秀人 (72) 発明者 平 博順 (72) 発明者

本電信電話株式会社内

本電信電話株式会社内

東京都千代田区大手町二丁目3番1号 本程值電話株式会社内 (72) 聲明者 中島 浩之 中国 勢 (72) 発明者

Œ

哲康 加品 (72) 発明者

東京都千代田区大手町二丁目3番1号 本電信電話株式会社内

東京都千代田区大手町二丁目3番1号 日

œ

•

本電信電話株式会社内

レロントページの結束

東京都千代田区大手町二丁目3番1号

原田 昭一 (72) 発明者

ш 東京都千代田区大手町二丁目3番1号

Fターム(参考) 58075 ND03 NK32 PP24 P002 P074 58091 AA11 AB06 CA02